AD-A238 832

ARI Research Note 91-72

Eliciting Knowledge From Military Experts: An Associative Network Approach

James E. McDonald

New Mexico State University

Field Unit at Fort Huachuca, Arizona
Julie A. Hopson, Chief

Systems Research Laboratory
Robin L. Keesee, Director

June 1991

91-06061

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

91

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

New Mexico State University

Technical review by

Michael J. Barnes

A-1

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approval for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 91-72
6a. NAME OF PERFORMING ORGANIZATION James E. McDonald 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute Field Unit at Fort Huachuca
6c. ADDRESS (City, State, and ZIP Code) Psychology Department New Mexico State University Las Cruces, NM 88003	7b. ADDRESS (City, State, and ZIP Code) ATTN: PERI-SA Fort Huachuca, AZ 85613-7000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-S	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAL03-86-D-0001, DO 2247-01
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 62785A PROJECT NO.: 790 TASK NO.: 1306 WORK UNIT ACCESSION NO.: C08

11. TITLE (Include Security Classification)
Eliciting Knowledge From Military Experts: An Associative Network Approach

12. PERSONAL AUTHOR(S)
McDonald, James E.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 90/08 TO 90/12	1991, June	32

16. SUPPLEMENTARY NOTATION
Julie A. Hopson, Contracting Officer's Representative

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Knowledge elicitation; Graphic analysis; Pathfinder; Paired comparisons; Psychophysical scaling; Associative networks

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

For this report, pathfinder and hierarchical cluster analyses were used to analyze knowledge elicited from military intelligence experts. Researchers derived associative networks for both individual and average data sets. Judges also graded the different networks. Results indicated that the networks were a reasonable representation of the military intelligence domain examined. In addition, the judges preferred the average network over their own networks. The techniques may provide a means to describe individual mental models in order to study differences in performance attributed to individual differences.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL David D. Burnstein	22b TELEPHONE (Include Area Code) (602) 538-4704	22c. OFFICE SYMBOL PERI-SA

DD Form 1473, JUN 86 Previous editions are obsolete.

ELICITING KNOWLEDGE FROM MILITARY EXPERTS: AN ASSOCIATIVE NETWORK APPROACH

CONTENTS

Page

LIST OF FIGURES

ELICITING KNOWLEDGE FROM MILITARY EXPERTS: AN ASSOCIATIVE NETWORK APPROACH

Introduction

There are numerous circumstances in which it is desirable to have an accurate representation of the knowledge experts use in making decisions. These include situations in which it is necessary to evaluate the decision-making processes of individuals or groups, to aid decision-making, or even "mimic" the decision-making processes of experts. Knowledge elicitation is very much an art, rather than a science, and the methodologies currently employed often produce mixed results. Eliciting the knowledge necessary to build "expert systems," for example, has proved to be a particularly difficult task. In fact, the problem of knowledge elicitation for building expert systems is so fundamental that it is often referred to as the knowledge-engineering bottleneck (Hayes-Roth, Waterman, & Lenat, 1983; Cooke & McDonald, 1986).

The two most popular techniques for eliciting knowledge are the interview and "thinking aloud" protocol analysis. Neither of these techniques is particularly effective. Interviews force experts to consciously introspect on their knowledge, then verbally express it in the form of rules. Both of these tasks are generally difficult--and there is no indication that experts fare any better than non-experts. The notion of inferring expert knowledge from behavior is appealing. Unfortunately, there are no formal procedures available for extracting knowledge directly from behavior, although statistical techniques may be useful in certain applications. The use of statistical techniques, however, requires large amounts of data and is therefore not applicable to most domains of expertise. In practice, protocols are often little more than structured interviews and, as such, suffer from the same limitations.

Because of the lack of formal (i.e., well-specified) procedures, the results of interviews and protocol analyses rely on, and are heavily influenced by, the domain knowledge of the knowledge engineer. This is obviously true when the domain expert and knowledge engineer are the same person, a common situation. In fact it is often said that such informal techniques work well only when the knowledge engineer becomes a domain expert--or the domain expert becomes a knowledge engineer.

Requiring the knowledge engineer to become a domain expert--or visa versa--is unsatisfactory because it is extremely demanding and time consuming. Furthermore, the results tend to be highly idiosyncratic and difficult for others to interpret or use effectively. The methodology presented attempts to address these problems by requiring little or no introspection by experts.

Only simple, discrete judgments--and correspondingly simple responses--are necessary. Furthermore, it consists of a set of well-specified steps which can be applied to a wide variety of domains of expertise. This minimizes the impact of knowledge-engineer subjectivity and does not require that the knowledge-engineer become a domain expert.

Looked at from another level of analysis, knowledge-transfer problems often result from basic differences in the nature of knowledge required for a particular application and the nature of expert knowledge. For example, domain experts may find it difficult to express their knowledge in the form of rules, a requirement of typical production-systems, because much of their knowledge doesn't consist of rules. It is even possible for experts to generate rules governing particular situations which bear little or no relationship to their actual behavior. Experts may be better at applying rules than novices, but rule-based reasoning is not the hallmark of expert performance. Experts excel at recognizing and classifying complex patterns of information in their domains of expertise and at building and using associations. This ability has been recognized in such diverse domains as chess (Chase & Simon, 1973), electronics (Egan & Schwartz, 1979), and computer programming (McKeithen, Reitman, Rueter, & Hirtle, 1981), among many others.

Much of what constitutes expert behavior might be characterized as intuitive, in that "...the individual has a sense of what is right or wrong, a sense of what is the appropriate or inappropriate response to make in a given set of circumstances, but is largely ignorant of the reasons for that mental state." (Reber, 1989, p. 232) In their influential book, Building Expert Systems, Hayes-Roth, Waterman, and Lenat (1983) note that "...knowledge does not appear in some precipitated form...[but as]...an unmined and unrefined substance" (p. 12). Furthermore, knowledge frequently consists of empirical associations, not rules. Much of expert behavior results from the ability to recognize complex patterns of information, not to effectively apply rules, and that the knowledge on which expert performance rests is seldom explicitly represented, but only implicit in the complex patterns of associations that experts develop through experience. From this perspective, the objective of knowledge elicitation is to infer knowledge structure from judgments about the relationships among domain concepts. This approach avoids the difficulties associated with having experts structure domain concepts or generate rules directly.

The methodology presented is a step-by-step procedure for acquiring and representing knowledge that relies on simple judgments of relatedness, minimizing introspection and speculation on the part of domain experts and knowledge engineers. It is designed to be more compatible with the nature of expertise, and does not require

experts to speculate on the way domain concepts are related or organize them within some prescribed framework. Once judgments are obtained, data scaling techniques are applied by the knowledge engineer to extract and represent the underlying structure.

The general methodology for eliciting and representing knowledge consists of: (1) identifying the important actions, objects, tasks, etc., in the target domain (hereafter referred to as domain concepts); (2) obtaining estimates of conceptual distance for all pairs of domain concepts, and (3) analyzing the obtained matrix of distances, or proximities, using scaling techniques (e.g., Pathfinder network analysis).

Identifying Domain Concepts

For some applications the process of identifying system concepts is deceptively straight-forward (e.g., UNIX commands. See McDonald and Schvaneveldt, 1988, for a discussion of this problem). Typically, however, the set of concepts is undefined or, at best, fuzzy. One technique is simply to have experts generate domain concepts in a relatively unconstrained fashion (e.g., listing important concepts). Other techniques include having experts generate scripts for domain-related tasks, list chapter titles, and subtitles for hypothetical books, or to extract critical ideas from interviews. McDonald, Dearholt, Paap, & Schvaneveldt (1986); and Anderson, McDonald, & Schvaneveldt, (1987) have also explored the use of event records, obtained from experts performing domain-relevant tasks and text analysis techniques for identifying domain concepts. For the present application the domain concepts consisted of a set of 42 items selected from the User Information Requirements Profile developed by ARI for assessing intelligence production (Burnstein, Fichtl, Landee-Thompson, & Thompson, 1990).

Obtaining Distance Estimates

Central to the methodology is the ability to obtain some measure of conceptual distance (or dissimilarity) among concepts. The pair-wise distances necessary for scaling analysis can be obtained from sources such as psychophysical judgments, frequency of co-occurrence, confusability, correlations, and temporal or spatial distance. A more complete discussion of these and other techniques is provided in McDonald and Schvaneveldt (1988).

Paired Comparison. The approach presented uses the method of paired comparison. In its simplest form, the method requires each expert to supply an estimate of relatedness for all $(n^2-n)/2$ pairs of concepts (i.e., the combination of n items taken two at a time). This technique offers several advantages. First, a great deal of information can be obtained from each expert.

Furthermore, each judgment is relatively simple and does not encourage judges to introspect on their knowledge, or to reflect on some overall organization for the set of concepts. The major problem with the paired-comparison technique is that it is inefficient and may not be suitable for many applications. For example, in order to obtain the distance estimates for the 42 items used in this study, each expert had to rate 861 pairs. Based on experience, it takes judges at least six seconds to rate a pair of items using interactive, computer-based, techniques. It probably takes twice that long using the method employed in the present study. Thus, rating 861 pairs would take at least three hours.

Although other techniques can be used for obtaining distance estimates from experts, such as sorting and repertory grid, each has its own problems. For the present study, the method of paired comparison offers the further advantage of minimizing the impact of prior domain taxonomies such as the User Information Requirements Profile on expert knowledge.

Psychophysical Scaling

Once the distance estimates have been obtained, they are analyzed using psychophysical scaling techniques. Generally there are two goals shared by most of these methods: (1) to reduce a large number of pair-wise distances to a simpler representation, and (2) to give some insight into the organization underlying the pair-wise distances. Techniques for representing distance estimates can be subdivided into two major classes, those that yield continuous models of mental structure and those that yield discrete models. Most of these methods take pair-wise distances as input and produce spatial or graph-theoretic representations as output.

Continuous Models. Continuous, or spatial, models represent entities as points in a multidimensional space. These models attempt to provide some global information about the entities represented, in that the dimensions of the space may correspond to abstract dimensions underlying the variations among the entities. Multidimensional scaling (MDS) is actually a general term used for a variety of specific techniques, all of which represent the scaled entities as points in multidimensional space (Kruskal, 1964, 1977; Shepard, 1962a, 1962b, 1963). The objective of these techniques is to find a spatial layout of the points that best corresponds to the given pair-wise distances.

Although MDS can represent large amounts of data in a form that is amenable to interpretation, it fails to capture some important aspects of the psychological organization of some domains. For example, if the domain consists of a heterogeneous collection of concepts, a spatial representation may necessitate a large number of

dimensions in order to capture the various ways in which the concepts are related. However, more than three dimensions are difficult to interpret and are certainly difficult to represent. There are some other problems that naturally follow from spatial representations. It is common for the psychological interpretation of the relations between two entities to be asymmetrical, as discussed by Tversky (1977). For example, Cuba seems to be more similar to Russia than Russia is to Cuba. Such asymmetries are inherently incompatible with spatial representations.

Another potential problem with MDS is that each of the distances plays an equal role in constraining the solution. Under some circumstances the equal weighting of the distances is a strength of the method since the entire data set of distances determines the representation. However, when MDS is used to scale psychological relatedness, equal weighting of the judgments may be inappropriate. In particular, it may be more important to focus on the relations among the most related concepts, to a large extent ignoring distances corresponding to "unrelated" judgments. Equal weighting of related and unrelated concepts can lead to distortions in the representation in which the points corresponding to related concepts are moved in the space in order to accommodate points that are not related. Because relations are represented solely by distance in MDS, this problem is difficult to resolve.

Discrete Models. Several psychometric methods derive from the mathematical theory of graphs (Carre, 1979; Christofides, 1975; Gibbons, 1985; Harary, 1969). A graph can be displayed by a diagram in which nodes are shown as points, and links are indicated by lines connecting appropriate pairs of points. In graph applications, each node represents an entity or concept, and the links represent pair-wise relations between entities. A wide variety of structures can be represented by graphs because a set of nodes can be connected by links in many possible ways.

Trees, graphs without cycles, are the basis of such psychometric methods as hierarchical cluster analysis (Johnson, 1967). weighted free trees (Cunningham, 1978), and additive similarity trees (Sattath & Tversky, 1977). Hierarchical cluster analysis provides a set of nested (hierarchical) groupings of entities which are meant to correspond to meaningful categories. Different hierarchical clustering methods use different definitions of the distance between a newly-formed category and other entities in categories. The single link method uses the minimum of the distances between the entities in a category and the entities in other categories. The complete link method uses the maximum distance. The value of hierarchical cluster analysis lies in its potential for revealing the underlying categorical structure for a set of entities. This method does have its problems, however, such as the necessity for clusters to be discrete, meaning each entity can only belong to one cluster.

Pathfinder. The representation technique investigated is most closely realized in an algorithm (Pathfinder) which is based on the idea that a link is present in an output network if and only if that link is the minimum weight path between nodes connected by the link in the (complete) network corresponding to the input distances (Schvaneveldt, Durso, & Dearholt, 1989). Links are given weights corresponding to the pair-wise distances given as input, and the weight of a path is a function of the weights of the links in the path. Pathfinder as described by Dearholt and Schvaneveldt (1990) depends on two parameters. "The first, the Minkowski r-metric, determines how distance between two nodes not directly linked is computed. The weight of a path with weights w_1, w_2,...w_k is:

$$W(P) = [\sum_{i=1}^{k} w_i^r]^{1/r}$$

"...The second is the q parameter, which is a limit on the number of links in the path examined in constructing a network. Its value determines the maximum number of links in paths in which the triangle inequalities are guaranteed to be satisfied in the resulting network." (p. 3) Several studies have shown that associative networks provide a good account of expert knowledge (Cooke, Durso, & Schvaneveldt, 1986; Goldsmith & Johnson, 1990).

The present study was intended to provide a basis for evaluating the associative network knowledge-elicitation methodology for various military intelligence applications.

Phase 1: Judging Relatedness

Method

Subjects. Two military intelligence experts participated in this study. Both had G2 experience.

Materials. Forty-two information items (see Appendix A) were selected from the User Information Requirements Profile. A list of 861 pairs of these information items was constructed, consisting of all possible pairs in the 42 items (n^2-n/2). Rating booklets were prepared, each consisting of instructions to the judges (see Appendix B), a list of the 42 items used in the study, a unique random sequence of the 861 information-item pairs, and computer scoring sheets for the ratings.

Procedure. Judges were instructed to examine the set of information items in order to familiarize themselves with the domain concepts and to allow them to "calibrate" their ratings. They were then to decide how related each of the 861 pairs of information-items were in the context of the entire set of items. Judges were encouraged to rely on experience and intuition in making their ratings, and not to spend too much time on any one judgment. Because of the large number of pairs, judges were instructed to proceed by first working through the list quickly, marking all of the pairs that were related. After the first pass through the list, they were to review the marked pairs, assigning relatedness values to them using a five-point scale in which "A" signified "Very Related" to "E" signified "Slightly Related." In an effort to reduce the tedium associated with marking the "bubbles" on the computer scoring sheets, judges were told that they did not have to mark pairs that were unrelated (i.e., "F"). Judges were asked to use the full range of the scale by assigning pairs of information items they judged to be somewhat related intermediate values (i.e., "B" through "E").

Results and Discussion

The data sheets from the two judges were scored and converted into electronic form. The data matrices were then reordered to correspond to the original information-item sequence. Ratings from the two judges were then averaged together. This resulted in three data sets: one for each expert (E1 and E2) and the average set (AVE) which reflected their combined judgments. Coherence measures were computed for each expert (E1 = .458, E2 = .680). This measure reflects how consistent judges are in assigning ratings (the higher the coherence measure, the more consistent the judge was in rating the concepts). High coherence measures may indicate the use of a well-formed conceptual model.

The three sets of data were separately submitted to Pathfinder analyses ($r=\infty, q=n-1$) and the resulting networks were graphically represented and compared. All three networks consisted of 42 nodes, corresponding to the 42 information-items, and a variable number of links, signifying strong associations among information-item nodes. The E1 network consisted of 137 links with a minimum link weight of 0 (corresponding to "A") and a maximum link weight of 2 (corresponding to "C"). The maximum link is the largest value required to link the 42 nodes. The E2 network contained 99 links, with a minimum link weight of 0 and a maximum link weight of 3. The AVE network contained an intermediate number of links (104), with a minimum link weight of 0 and a maximum link weight of 2. This result is surprising because average data usually results in a network which has fewer links than any of the contributing individuals.

Although the resulting networks appeared sensible, implying that subjects were correctly basing their judgments on their conceptual models of military intelligence, they were unusually

dense (i.e., they contained a very large number of links for the number of nodes) even though the Pathfinder parameters used in these analyses ($r=\infty, q=n-1$) produce the sparsest networks for a given set of data. Compared to typical results in past research, the rating distributions for the two experts reveal a tendency to evaluate a large number of concepts pairs as "Very related," which results in a large number of tied values (see Figure 1). This was particularly true of E1. Because Pathfinder includes all links when there are tied distances, the large number of links can be attributed to this failure to discriminate among highly related concepts. Although the judges had been instructed to use the entire scale, the instructions appear to have been inadequate in encouraging them to make the fine discriminations necessary for clear structures to be revealed.

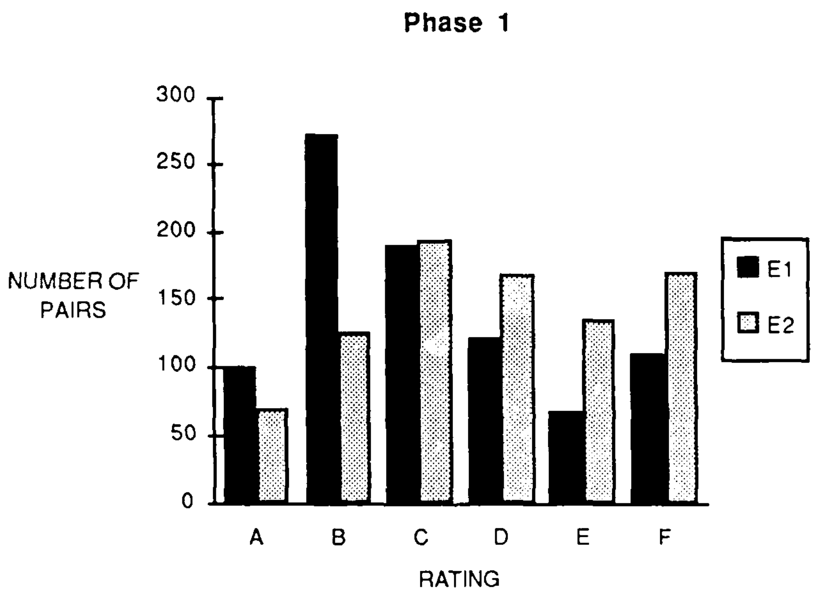


Figure 1. Frequency distributions for the relatedness ratings obtained from the two military intelligence experts during Phase 1.

Phase 2: Refining the Relatedness Judgments

Method

Subjects. The two military intelligence experts from Phase 1 participated in Phase 2.

Materials. A subset of the original concept pairs was identified by combining the 137 most closely-linked pairs from each experts' Pathfinder networks. Since 33 pairs were common to both experts, this procedure resulted in a list containing the 241 most related concept pairs (137+137-33).

Procedure. Instructions to judges were similar to those of Phase 1 (see Appendix C). However, special emphasis was placed on using the full range of the scale. The intent was to have judges make finer discriminations among the set of concept pairs than they had made in Phase 1.

Results and Discussion

The attempt to increase discriminability among the most related concept pairs appears to have been successful (see Figure 2). Both experts used the entire range of the rating scale for the 241 concept pairs, confirming the hypothesis that they were capable of making finer distinctions than was evidenced in Phase 1.

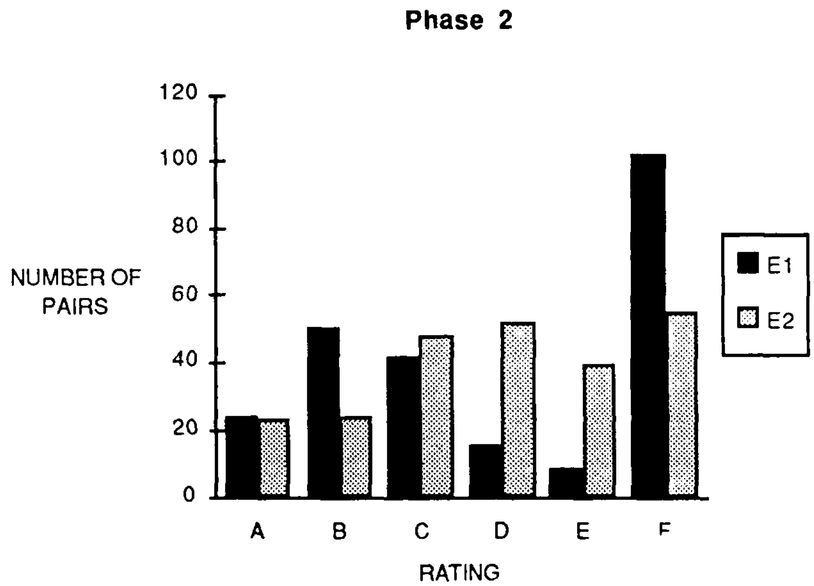


Figure 2. Frequency distributions for the relatedness ratings obtained from the two military intelligence experts during Phase 2.

Distance matrices were constructed for each of the experts by assigning their 241 ratings to the appropriate concept pairs in individual data matrices. Because they were not rated during Phase 2, the remaining 620 pairs (861-241) were given values corresponding to one more than "Unrelated" (i.e., 6). The data from the two experts (E1B and E2B) were also averaged to produce an AVEB matrix.

The three matrices were each submitted to Pathfinder analysis $(r=\infty, q=n-$ and graphic representations were produced using the MacKnot™ software package. As in Phase 1, all three networks consisted of 42 nodes, corresponding to the 42 information-items, and a variable number of links. The E1B network consisted of 97 links with a minimum link weight of 0 and a maximum link weight of 5. It was necessary to include links rated F to insure a completely

connected network. The E2B network contained 78 links, with a minimum link weight of 0 and a maximum link weight of 4. The AVEB network contained 63 links, with a minimum link weight of 0 and a maximum link weight of 4. There were 39.5% fewer links in the average network from Phase 2, compared to Phase 1. These networks were relatively sparse and sufficiently interpretable to serve as stimuli for the third and final phase.

Phase 3: Network Evaluation

Method

Subjects. The same two military intelligence experts from Phases 1 and 2 participated in Phase 3.

Materials. The three networks produced in Phase 2 (E1B, E2B, and AVEB) served as stimuli for the third phase (see Appendix D). In addition, a fourth network was generated from the E1B data by limiting the number of links to 78 (the same as the E2B network). The purpose of including this additional network was to control for the tendency to judge sparse networks more favorably than dense networks.

Procedure. Judges were instructed to examine the four networks, paying attention primarily to link structure. They were told to avoid giving much importance to the layout (i.e., placement) of the nodes because in the layout procedure used, node placement was largely determined by link structure. Judges were to focus on connections (i.e., links). They were told to examine each network, marking any links they thought particularly good or bad. Good links were defined as those that highlighted particularly strong associations in this domain, whereas bad links were those in which two concepts were linked in the network although they had only a limited or trivial association in the domain. Judges were also encouraged to note any "missing" links.

Judges were encouraged to write notes on the networks. Their final task was to rank-order the four networks from good to bad, using the above criteria, and to report the numbers corresponding to their selections. In rank-ordering the networks, judges were to consider the extent to which each network captured the structure of knowledge in the military intelligence domain and assign the network an overall "grade" from "A" to "F."

Results and Discussion

Based on the comments received from the judges, it is not clear whether or not the two experts evaluated the networks independently. With that in mind, their evaluations of the networks corresponded to expectations. The network based on the average data (AVEB) was given the highest grade ("B"). The E2B network was judged the next best and given a grade of "C." The two networks based on the E1B

data were judged the worst by both experts and given identical grades of "D" (original 98 link network) and "F" (modified 78 link network). The poor showing of the modified network was expected, since reducing the number of links meant that several of the concepts were disconnected.

One particular interesting result from Phase 3 was that both experts attempted to identify groupings of concepts in the networks, even though they were not instructed to do so. This technique is used often to simultaneously display information from Pathfinder and cluster analyses.

Hierarchical Cluster Analyses. In an effort to further clarify the network analyses, both single-line (minimum) and complete-link (maximum) hierarchical cluster analyses were performed on the AVEB data. Because of its tendency to quickly agglomerate items, however, the single-link solution was not particularly informative. As a result, only the complete-link analyses was performed on the individual data. The results of the complete-link analysis were used to guide manual restructuring of the network layout for the AVEB and are shown in Figure 3 (A version of Figure 3 with clusters is included in Appendix E. Hierarchical cluster analyses are shown in Appendix F).

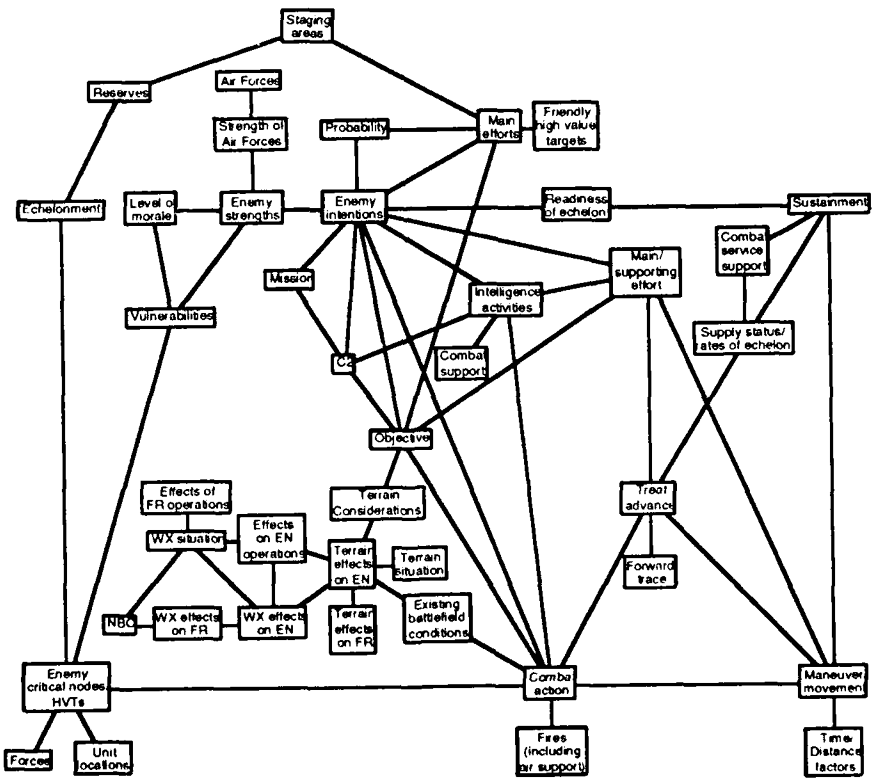


Figure 3. Final Pathfinder network layout derived from AVEB data and restructured according to complete-link cluster analysis (see Appendix E for more details).

Graph-Theoretic Distance Estimates. In addition to the preceding analyses, a network was produced by averaging the graph-theoretic distances derived from the experts Pathfinder networks (Schvaneveldt, 1990). The graph-theoretic distance is defined as the minimum number of links connecting two nodes in a graph (in this case the expert's Pathfinder networks). This is a special case of distance in a network in which the weight on each link is considered to be one. The averaging operation presupposes that the data have ratio properties, and rating data can seldom be assumed to have more than ordinal properties. Link distances in the Pathfinder network, however, are true distances with ratio properties and the purpose of this analyses was to overcome any anomalies that may have arisen from averaging the rating data directly. A secondary purpose was to provide a representation suitable for displaying links common to both experts Pathfinder solution. The Pathfinder analyses of the average graph distance data contained 59 links, of which 32 were common to both expert's networks (see Appendix G).

Conclusions

The associative network approach to knowledge elicitation is promising. Although only two experts were available as subjects, a reasonably coherent representation of a portion of the military intelligence domain was obtained. Although the best network representation (from the average data) was rated only "good" by the two military experts, their comments suggest that the networks were "intriguing" and that it captured many subtle distinctions. It is likely that the final network representation, restructured to accentuate complete-link clusters of concepts, would be judged even more favorably.

Although associative networks have been successfully used to classify individuals as members of particular groups (Schvaneveldt & Goldsmith, 1986; Goldsmith & Johnson, 1990), it is a far more difficult task to contrast the knowledge structures of individuals. The most that can be attempted from these results is to speculate about the strength of the two experts' conceptual models. From the fact that Expert 2 was better able to make fine discriminations among the highly related domain concepts, both in Phase 1 and Phase 2, it would appear that he has a more coherent conceptual model than Expert 1. This does not in and of itself imply that Expert 2 has a better or more accurate model, but only a more structured model (also supported by the individual's complete link cluster analysis). The fact that both experts favored the conceptual model of Expert 2 more than that derived from the ratings of Expert 1 does suggest that the second experts's model better captures relationships in this domain. However, the best model resulted from the combined data for Expert 1 and Expert 2, indicating that each expert had something unique to contribute.

Recommendations for Future Work

Several recommendations can be made based on the results of this study. First, those concepts that clustered late in the complete-link analysis should be closely examined (Appendix E). Late clustering of concepts may result from disagreement about the meaning of the concepts, or it may mean that the concepts are not very important for this domain. The complete-link cluster analysis could also be used to make a direct comparison with the categories of information in the User Information Requirements Profile. This comparison could potentially highlight differences between the top-down approach used to structure the User Information Requirements Profile, and the context-free, bottom-up approach used in these analyses. The final Pathfinder network can be used in a similar way to indicate concepts that are strongly related to more than one category.

Finally, the complete-link cluster analysis may provide a means to determine how different experts approach a problem and how they might be expected to sequentially use information to reach different solutions. The Pathfinder and hierarchical cluster techniques may have significant potential for the study of how individual differences contribute to decision making and problem solving performance.

References

Anderson, M. P., McDonald, J. E., & Schvaneveldt, R. W. (1987). Empirical user modeling: Command usage and task analyses for deriving models of users. Proceedings of the Human Factors Society 31st Annual Meetings (pp. 41-45).

Carre, B. (1979). Graphs and Networks. Oxford: Clarendon Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

Christofides, N. (1975). Graph Theory: An Algorithmic Approach. New York: Academic Press.

Cooke, N. M., & McDonald, J. E. (1986). A formal methodology for acquiring and representing expert knowledge. Proceedings of the IEEE: Special Issue on Knowledge Representation, 74, 1422 -1430.

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 538-549.

Cunningham, J. P. (1978). Free trees and bidirectional trees as representations of psychological distance. Journal of Mathematical Psychology, 17, 165-188.

Dearholt, D. W. & Schvaneveldt, R. W. (1990). Properties of pathfinder networks. R. W. Schvaneveldt (Ed.) Pathfinder Associative Networks: Studies in knowledge organization, 1-30. New Jersey: Ablex Publishing Corporation.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. Memory and Cognition, 7, 149-158.

Gibbons, A. (1985). Algorithmic Graph Theory. Cambridge: Cambridge University Press.

Goldsmith, T. E., & Johnson, P. J. (1990). A structural assessment of classroom learning. In R. Schvaneveldt (Ed.), Pathfinder associative networks: Studies in knowledge organization (pp. 241-254). Norwood, NJ: Ablex.

Harary, F. (1969). Graph Theory. Reading, MA: Addison-Wesley.

Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). Building Expert Systems. Reading, MA: Addison-Wesley Publishing Company, Inc.

Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32, 241-254.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. Psychometrika, 29, 115-129.

Kruskal, J. B. (1977). Multidimensional scaling and other methods for discovering structure. In Enslein, Ralston, and Wilf (Eds.), Statistical Methods for Digital Computers. New York: Wiley.

McDonald, J. E., & Schvaneveldt, R. W. (1988). The application of user knowledge to interface design. In R. Guindon (Ed.), Cognitive science and its applications for human-computer interaction (pp. 289-338). Hillsdale, NJ: Erlbaum.

McDonald, J. E., Dearholt, D. W., Paap, K. R., & Schvaneveldt, R. W. (1986). A formal interface design methodology based on user knowledge. CHI '86 (pp. 285-290).

McKeithen, K. B., Reitman, J. R., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. Cognitive Psychology, 13, 307-325.

Reber, A. S. (1989). Implicit learning and tacit knowledge. Journal of Experimental Psychology: General, 118, 219-235.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. Psychometrika, 42, 319-345.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory, Vol 24 (pp. 249-284). New York: Academic Press.

Schvaneveldt, R. W., & Goldsmith, T. E. (1986). A model of air combat decisions. In E. Hollnagel, G. Mancini, & D. Woods (Eds.), Intelligent decision support in process environments (pp. 395-405). Berlin: Springer-Verlag.

Shepard, R. N. (1962a). Analysis of proximities: Multi -dimensional scaling with an unknown distance function. I. Psychometrika, 27, 125-140.

Shepard, R. N. (1962b). Analysis of proximities: Multi -dimensional scaling with an unknown distance function. II. Psychometrika, 27, 219-246.

Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. Human Factors, 5, 33-48.

Shepard. R. N. (1974). Representation of structure in similarity data: Problems and prospects. Psychometrika, 39, 373-421.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

APPENDIX A: THE FORTY-TWO DOMAIN CONCEPTS

1. Air Forces
2. C2
3. Combat action
4. Combat service support
5. Combat support
6. Echelonment
7. Effects of FR operations
8. Effects on EN operations
9. Enemy critical nodes/HVTs
10. Enemy intentions
11. Enemy strengths
12. Existing battlefield conditions
13. Fires (including air support)
14. Forces
15. Forward trace
16. Friendly high value targets
17. Intelligence activities
18. Level of morale
19. Main efforts
20. Main/supporting effort
21. Maneuver/movement
22. Mission
23. NBC
24. Objective
25. Probability
26. Readiness of echelon
27. Reserves
28. Staging areas
29. Strength of Air Forces
30. Supply status/rates of echelon
31. Sustainment
32. Terrain Considerations
33. Terrain effects on EN
34. Terrain effects on FR
35. Terrain situation
36. Time/Distance factors
37. Treat advance
38. Unit locations
39. Vulnerabilities
40. WX effects on EN
41. WX effects on FR
42. WX situation

APPENDIX B
PHASE 1 INSTRUCTIONS

A	B	C	D	E	F
very related		moderately related		slightly related	

Instructions to Judges

The purpose of this study is to discover the relationships between information items. You are being asked to participate as an expert in military intelligence. Your task will be to judge the relatedness of pairs of information items in terms of how related they are in your experience.

In addition to these instructions, you will be provided with a number of pages containing lists of information-item pairs. You are to decide how related each pair of information-item is in the context of the entire set of items (shown below). You should rely on your experience and intuition in making these judgments, rather than a deep analysis of the items. Because of the large number of pairs, we recommend that you proceed by first working through the list quickly, marking all of the pairs that you think are related (e.g., circle the pair number on the lists). After the first pass, you should go back to the marked pairs and assign relatedness values to them according to the scale shown at the top of every page. Once you have decided on the relatedness value for a pair, please fill in the corresponding circle on the data sheet (i.e., "A" through "E"). You *do not* have to mark the pairs that you judge unrelated (i.e., "F"). This procedure should save you considerable time, since many pairs will seem relatively unrelated within the context of military intelligence.

Please try to use the full range of the scale. If you think two information items are somewhat related, then you should assign them an intermediate value on the scale (i.e., "B" through "E"). Please don't spend too much time on any one decision. If you can't easily think of a way in which the two items are related, then they probably are *not related* in the context of military intelligence.

APPENDIX C
PHASE 2 INSTRUCTIONS

A	B	C	D	E	F
very related		moderately related		slightly related	

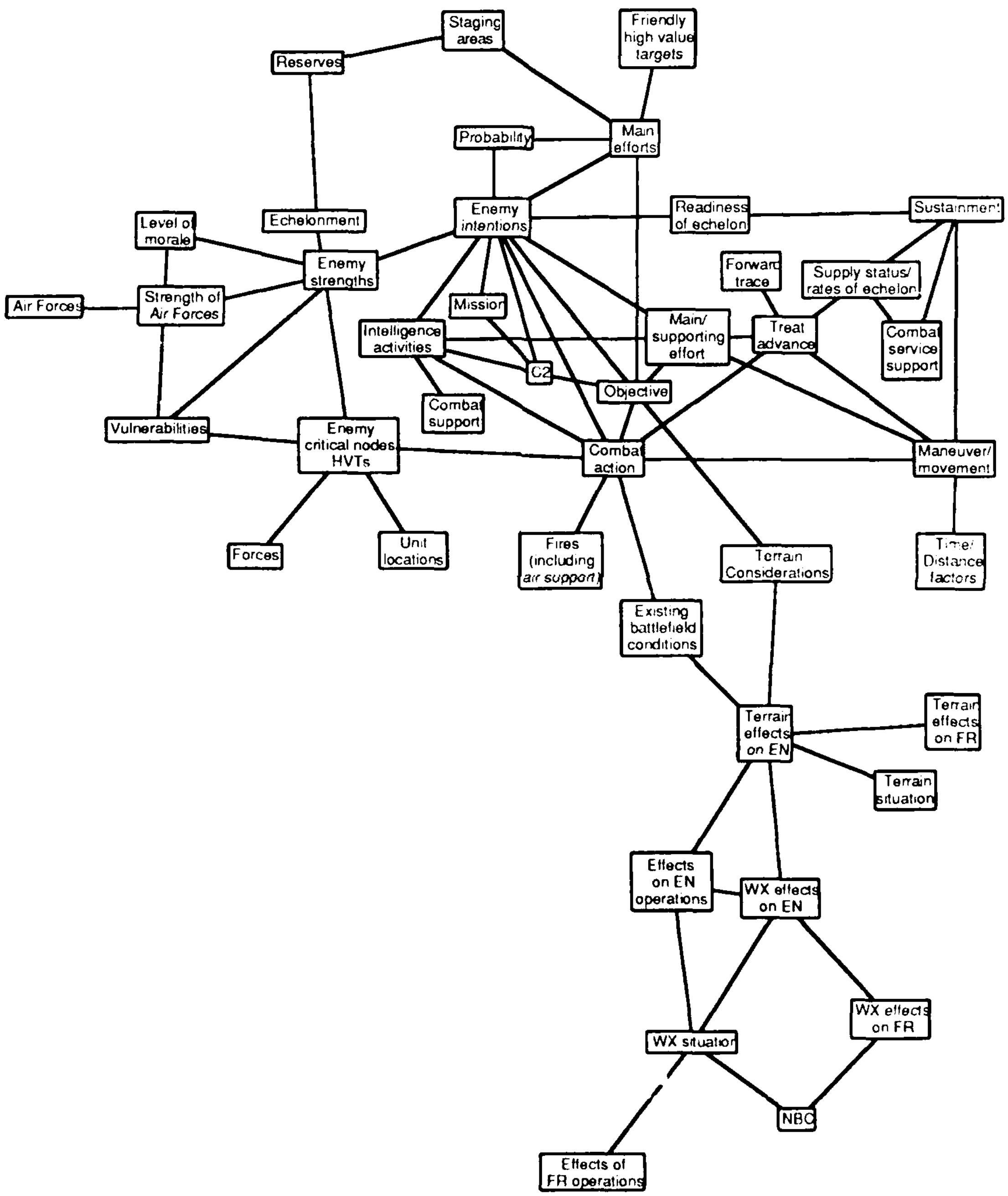
Instructions to Judges

The purpose of this study is to discover the relationships between information items. You are being asked to participate as an expert in military intelligence. Your task will be to judge the relatedness of pairs of information items in terms of how related they are in your experience.

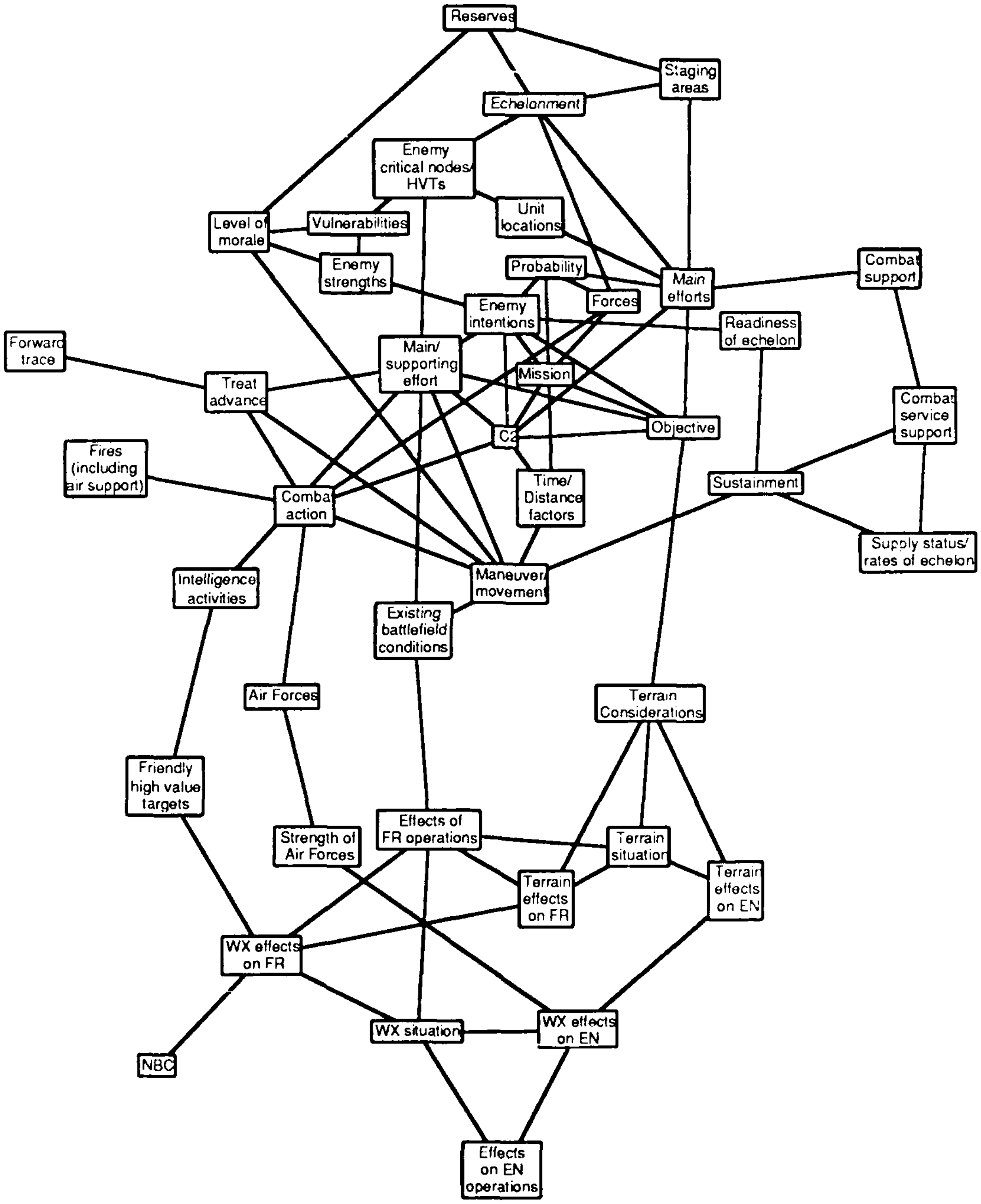
In addition to these instructions, you will be provided with several pages containing lists of information-item pairs. You are to decide how related each pair of information-items is in the context of the entire set of items (shown below). You should rely on your experience and intuition in making these judgments, rather than a deep analysis of the items. You should proceed by first working through the list quickly, marking all of the pairs that you think are related (e.g., circle the pair number on the lists). After the first pass, you should go back to the marked pairs and assign relatedness values to them according to the scale shown at the top of every page. Once you have decided on the relatedness value for a pair, fill in the corresponding circle on the data sheet (i.e., "A" through "E"). You need **not** mark the pairs that you judge unrelated (i.e., "F"). This should save you some time, since many pairs will seem relatively unrelated within the context of military intelligence.

Please try to use the full range of the scale. If you think two information items are somewhat related, then you should assign them an intermediate value on the scale (i.e., "B" through "E"), rather than marking all related pairs as "A.". You shouldn't spend too much time on any one decision. If you can't easily think of a way in which the two items are related, then they probably are *not very related* in the context of military intelligence.

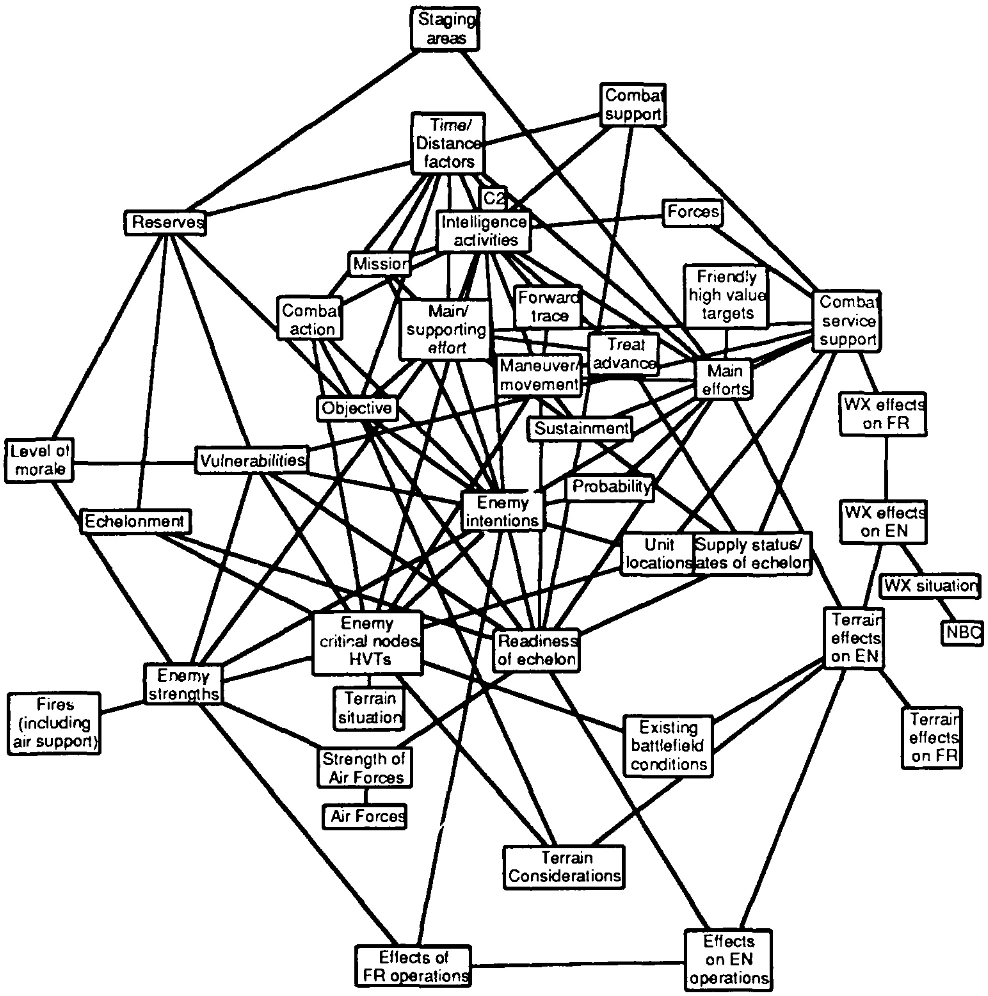
APPENDIX D
PFNETS USED IN PHASE 3



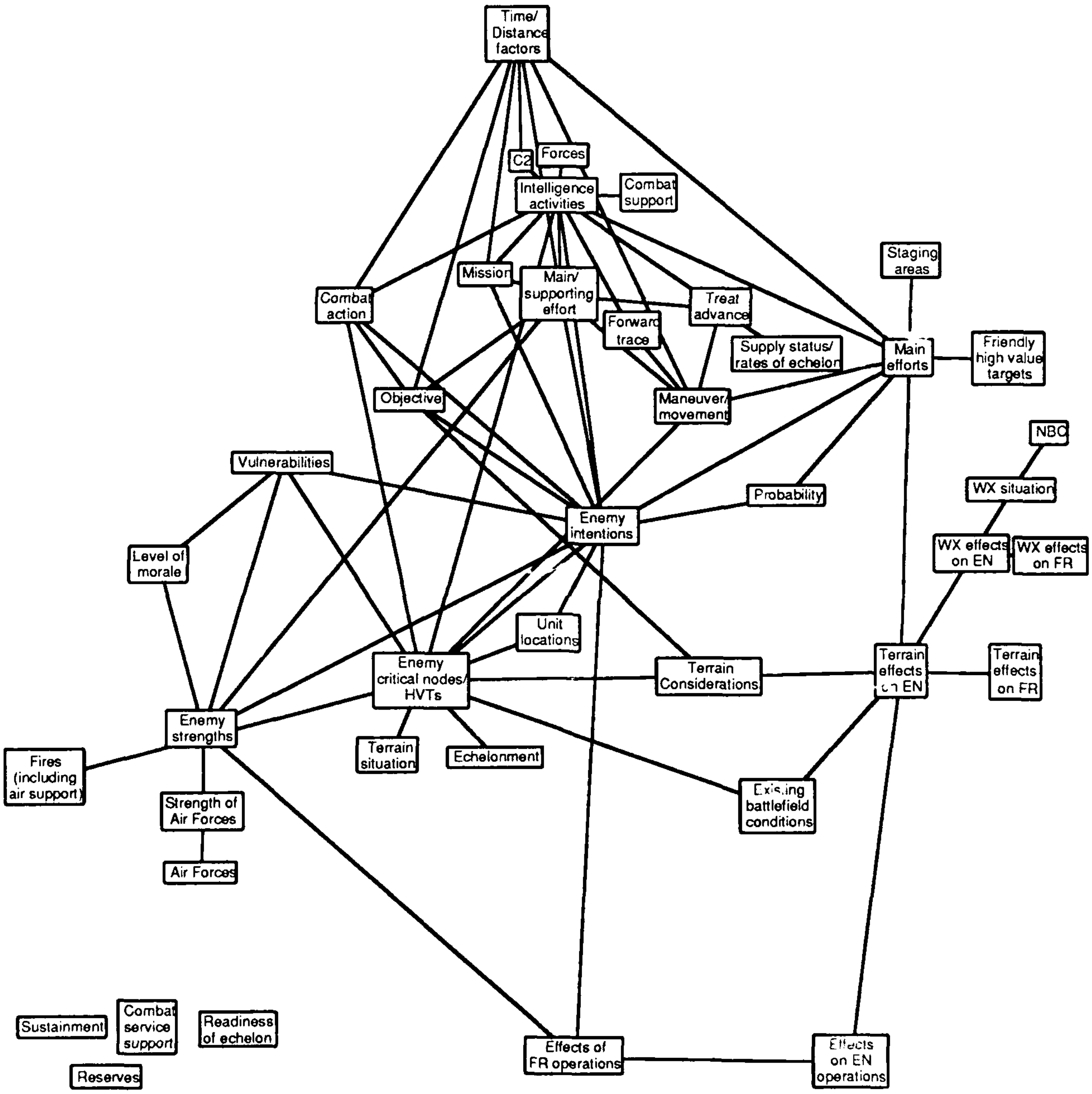
Network Based on Average Data (AVEB)



Subject E2B Network

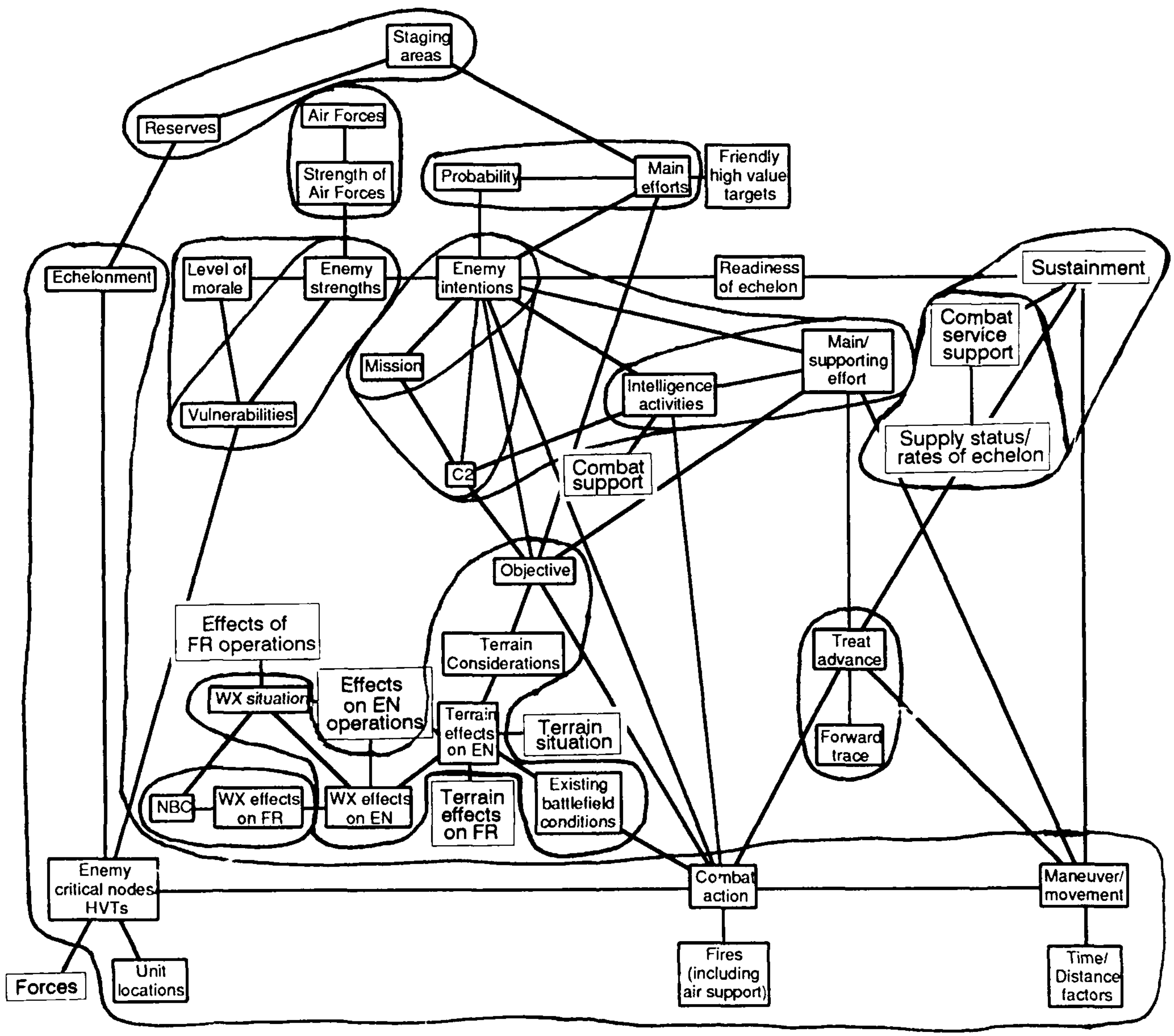


Subject E1B Network



Derived Network of Modified 78 Links

APPENDIX E:
AVEB PATHFINDER NETWORK[2]



Complete-link cluster analysis indicated using enclosing boundary lines wherever possible.

APPENDIX F:
HIERARCHICAL CLUSTER ANALYSES

Equivalence Class	Number of Items	Item Numbers
0	2	10 22
0	2	11 39
0.5	10	10 22 11 39 24 32 33 12 40 42
1	15	8 10 22 11 39 24 32 33 12 40 42 9 19 25 35
1	2	15 37
1	2	17 20
1.5	2	1 29
1.5	28	2 8 10 22 11 39 24 32 33 12 40 42 9 19 25 35 17 20 3 6 15 37 21 23 34 38 13 36
2	32	2 8 10 22 11 39 24 32 33 12 40 42 9 19 25 35 17 20 3 6 15 37 21 23 34 38 13 36 7 14 28 41
2.5	37	1 29 2 8 10 22 11 39 24 32 33 12 40 42 9 19 25 35 17 20 3 6 15 37 21 23 34 38 13 36 7 14 28 41 5 16 31
2.5	2	4 30
3	40	1 29 2 8 10 22 11 39 24 32 33 12 40 42 9 19 25 35 17 20 3 6 15 37 21 23 34 38 13 36 7 14 28 41 5 16 31 4 30 18
3.5	41	1 29 2 8 10 22 11 39 24 32 33 12 40 42 9 19 25 35 17 20 3 6 15 37 21 23 34 38 13 36 7 14 28 41 5 16 31 4 30 18 27
4	42	1 29 2 8 10 22 11 39 24 32 33 12 40 42 9 19 25 35 17 20 3 6 15 37 21 23 34 38 13 36 7 14 28 41 5 16 31 4 30 18 27 26

Single-link (minimum) cluster analysis (AVEB)

Equivalence Class	Number of Items	Item Numbers
0	2	11 39
0.5	6	12 33 32 40 24 42
1	2	15 37
1	2	17 20
1	2	19 25
1.5	2	1 29
1.5	3	2 10 22
1.5	7	3 9 13 21 6 36 38
2	2	23 41
2.5	2	4 30
2.5	4	7 8 34 35
3	3	4 30 31
3	3	11 39 18
3.5	5	2 10 22 17 20
3.5	2	27 28
4	5	4 30 31 5 14
6	42	1 29 2 10 22 17 20 3 9 13 21 6 36 38 4 30 31 5 14 7 8 34 35 11 39 18 12 33 32 40 24 42 15 37 16 19 25 23 41 26 27 28

Complete-link (maximum) cluster analysis (AVEB)

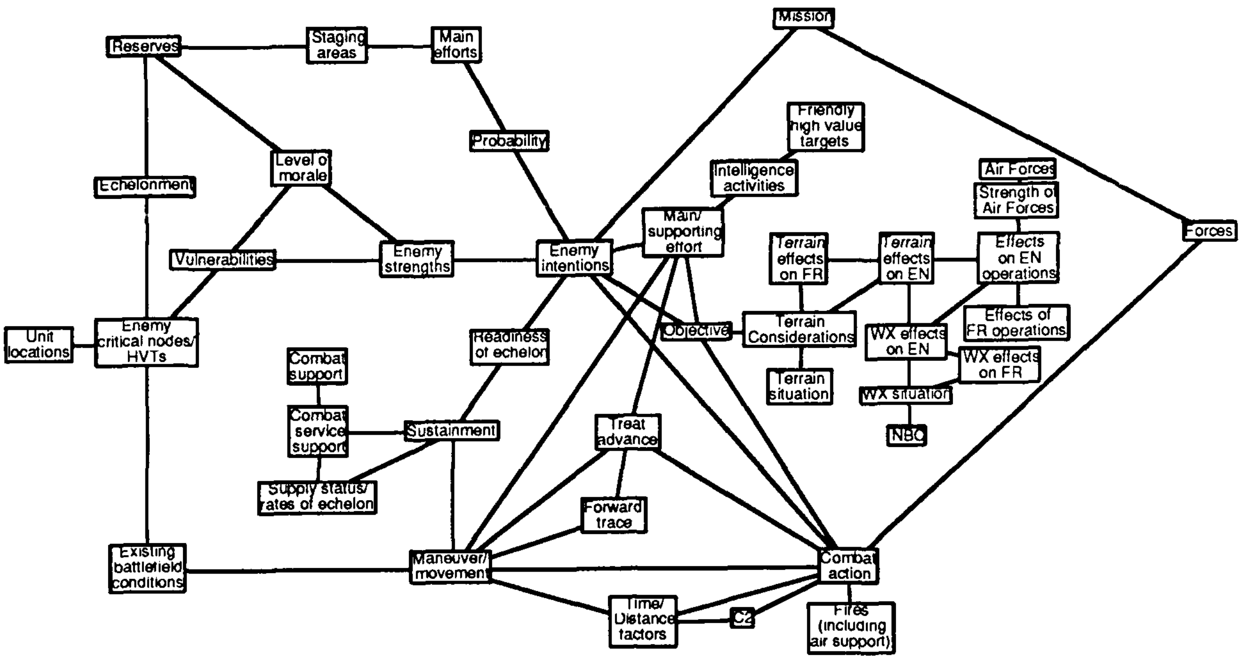
Equivalence Class	Number of Items	Item Numbers
0	16	2 17 3 5 9 10 14 15 19 22 37 11 12 24 33 39
1	2	20 21
1	2	40 42
2	5	7 8 38 40 42
2	3	20 21 36
3	2	1 29
3	2	32 35
4	2	23 41
5	12	4 30 31 13 26 6 27 28 34 18 32 35
6	42	1 29 2 17 3 5 9 10 14 15 19 22 37 11 12 24 33 39 4 30 31 13 26 6 27 28 34 18 32 35 7 8 38 40 42 16 20 21 36 23 41 25

Expert 1 (E1B data).

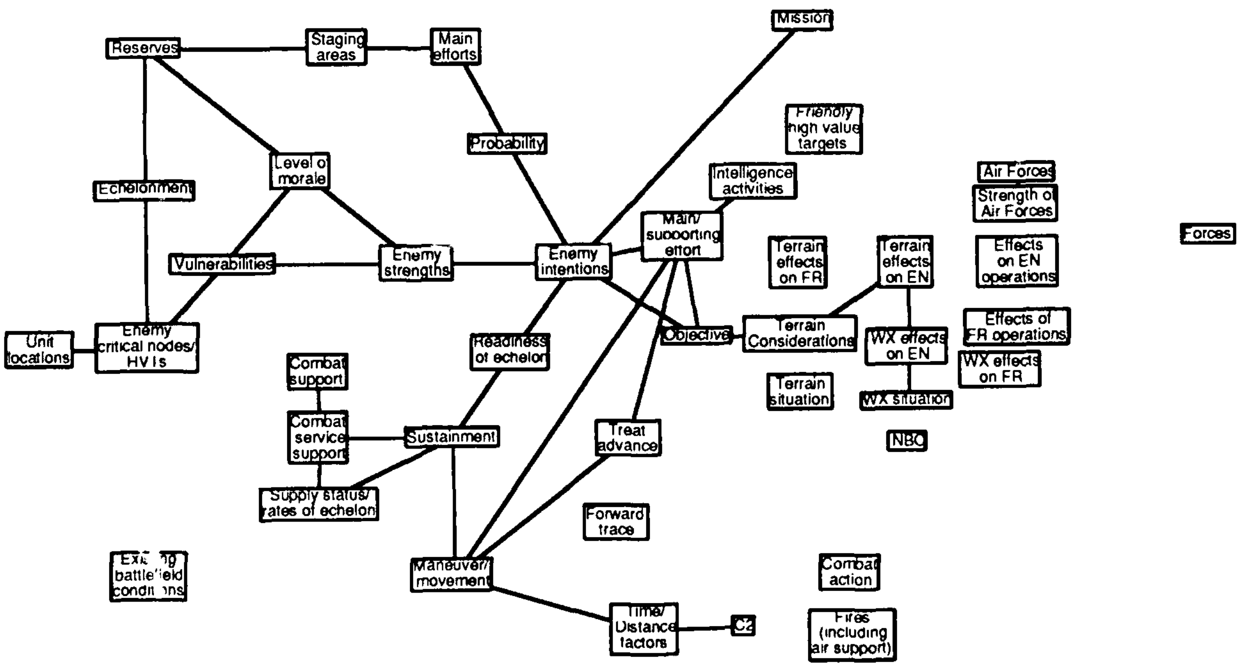
Equivalence Class	Number of Items	Item Numbers
0	2	1 29
0	2	4 30
0	12	7 12 34 35 41 42 8 23 32 33 40 24
0	2	10 22
0	2	11 39
0	2	21 31
1	2	3 13
1	2	6 14
1	2	15 37
1	2	17 20
1	4	19 25 38 36
2	3	2 10 22
2	2	27 28
3	5	3 13 9 6 14
3	3	4 30 5
4	5	2 10 22 17 20
4	3	11 39 18
5	3	21 31 26
6	42	1 29 2 10 22 17 20 3 13 9 6 14 4 30 5 7 12 34 35 41 42 8 23 32 33 40 24 11 39 18 15 37 16 19 25 38 36 21 31 26 27 28

Expert 2 (E2B data).

APPENDIX G:
PATHFINDER NETWORKS BASE ON GRAPH DISTANCES



Average Graph Distance Network



Average Graph Distance Network Showing Links in Common